DOCUMENT RESUME

ED 286 900 TM 870 530

AUTHOR Mandeville, Garrett K.
TITLE The Stability of School Effectiveness Indices across Years.
PUB DATE Apr 87
NOTE 26p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Washington, DC, April 21-23, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Tests; *Basic Skills; Correlation; Elementary Education; Longitudinal Studies; Mathematics Achievement; Multiple Regression Analysis; *Outcomes of Education; Predictive Measurement; Reading Achievement; *Reliability; *School Effectiveness; Socioeconomic Status; Statistical Studies; Trend Analysis
IDENTIFIERS Cognitive Skills Assessment Battery; Comprehensive Tests of Basic Skills; Effective Schools Research; *School Effectiveness Indices; *South Carolina Basic Skills Assessment Program; South Carolina Statewide Testing Program

ABSTRACT

School effectiveness indices (SEIs) based on residuals from regressing test performance in reading and mathematics onto prior-year test performance and a socioeconomic status measure (percentage eligible for a subsidized lunch program) were obtained for two consecutive years for 431 South Carolina elementary schools. The analysis involved school means based on matched student data, and was restricted to schools containing grades 1 through 4. The resulting SEIs were found to be reasonably stable across years, the correlations ranging from .34 to .66 depending on grade level and subject area. Reading-mathematics composites and cross-grade weighted averages were no more stable than grade-specific SEIs. Reading and mathematics SEIs, which were of roughly equal stability, decreased from grade 1 to grade 3 where the outcomes were measured by a local criterion-referenced test, the Basic Skills Assessment Program. For grade 4, where the norm-referenced Comprehensive Tests of Basic Skills were used, mathematics SEIs appeared to be more stable than those assessing reading. The relative stability of SEIs based on the same and different groups of students at the same and different grade levels suggests that grade-within-school effects dominate whatever global school effects operate in elementary schools. (Author/LPG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED286900

The Stability of School Effectiveness Indices Across Years

Garrett K. Mandeville
University of South Carolina

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. K. Mandeville

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running head: STABILITY OF SEIS ACROSS YEARS

Paper presented at the annual convention of the National Council on Measurement in Education, Washington D. C., April 1987

TM 870530

BEST COPY AVAILABLE

Abstract

School effectiveness indices (SEIs) based on residuals from regressing test performance in reading and mathematics onto prior year test performance and an SES measure were obtained for two consecutive years for a large sample of elementary schools. The analysis involved school means based on matched student data and was restricted to schools containing grades one through four. The resulting SEIs were found to be reasonably stable across years, the correlations ranging from .34 to .66 depending upon grade level and subject area. Reading and math SEIs, which were of roughly equal stability, decreased from grade 1 to grade 3 where the outcomes were measured by a local CRT. For grade 4, where an NRT was used, math SEIs appeared to be more stable than those assessing reading. The relative stability of SEIs based on the same and different groups of students at the same and different grade levels suggests that grade-within-school effects dominate whatever global school effects operate in elementary schools.

For a number of years, researchers have attempted to measure "schools performance." Most frequently quantitative indicators of student performance have been used and the focus has been on school accountability, "school effectiveness", and the more recent efforts to monitor schools and possibly award the ones whose students have exhibited exceptional achievement. Although most researchers admit that such attempts to measure school performance should be based on a variety of outcome measures, student test scores are still the primary or sole type of data utilized in most cases.

The recent impetus in this area is related to state- and district-level programs to monitor school performance on the basis of student achievement and other types of data. These programs often lead to the recognition of high performing schools and, in a few cases, monetary awards to the schools and/or their personnel (see, e. g., Wynne, 1984). According to Bandy (1985) six states have what are called "performance indicator" programs and three of them --California, Florida, and South Carolina -- have school award programs in which test scores are a major factor in the determination of awardees. Similar programs have been in operation in some of the larger school districts in the country.

This paper will focus on the year-to-year stability of "school effectiveness indices" (SEIs) for a large sample of elementary schools. SEIs will be operationalized in a manner which is consistent with most recommendations and appears to be growing in popularity. The SEIs will be based solely on achievement test data, however. Since cross-grade SEIs have been demonstrated to be

only weakly related, it is important to provide a more definitive assessment of cross-year stability in order to further address the issue of whether schools can be reliably measured on how they affect the achievement of their students.

It could be argued that if SEIs are unstable year-to-year, this would have important ramifications for researchers who use them to identify effective schools for further study. Good and Brophy (1986) liken this to studying a Super Bowl winner the year following their national championship when the odds are strong that the team will not repeat its exemplary performance. Although this research clearly has applicability to those involved in the identification and study of effective schools, this issue will not be dealt with here. To this writer, the identification of a few truly effective schools is a conceptually different task than the one confronting those charged with the responsibility of ranking essentially all of the schools in a district or state.

The counter argument is that year-to-year stability of SEIs is not relevant to the task of monitoring the progress of schools for the purpose of recognition or rewards. For that matter, it may be politically more acceptable if SEIs are unstable over years than if they are stable since then more schools will be honored. As educational researchers, however, we have no right to accept this argument. If we view our task as one of promoting school improvement, then information on the stability of SEIs, and the factors which affect that stability, is important.

Historical Background and Empirical Evidence

An Early Accountability Model

One of the early accountability models was a regression-based method developed by Dyer and his colleagues (Dyer, Linn & Patton, 1969; Dyer, 1970). In Dyer's method student outcomes (output variables) are regressed onto student inputs and "hard-to-change" variables, with the residuals serving as the SEIs. An important characteristic of Dyer's system is the use of longitudinal data, the justification being that, "the only fair index of school effectiveness is one that rests on input-output data obtained only on those pupils with whom the school staff has been in continuous contact over a specified period of months or years." (Dyer, 1970, p. 208). Results presented by Hilton and Patrick (1970) demonstrated differences between school aggregate indices based on matched longitudinal, unmatched longitudinal, and cross-sectional data. Related results in Dyer et al. (1969) also provided empirical support for this position. Because of the compelling nature of Dyer's arguments favoring the use of longitudinal data, the empirical studies included in the remainder of this review will be limited to those satisfying this restriction.

A related issue involves the unit of analysis to include in the regression analysis. Dyer et al. (1969) found that the residuals from a student level regression analysis aggregated to the school level were highly correlated with the residuals from an analysis involving school means (median r=.93). Furthermore, they found that the student level analysis produced SEIs which were slightly more stable. On the other hand, O'Connor (1972) has argued that a school level regression analysis is more appropriate because the aggregated residuals from a student level analysis need

not be uncorrelated with either input or predicted output variables. For this reason and the support provided by Frechtling (1982) and Purkey and Smith (1982) regression of school means was selected as the analytic strategy for this investigation.

Effective Schools Research

Researchers interested in developing evidence to refute the implications of the Coleman report (Coleman, et al., 1966) -- that schools made little or no difference -- attempted to identify a few exemplary schools which clearly did make a difference. Close scrutiny of these schools produced the research evidence which has served as the basis for a pervasive school improvement thrust in this country.

A large number of different approaches to the identification of these effective schools have been presented in the literature and Rowan, Bossart and Dwyer (1983) grouped them into four general methodological areas as follows: (1) using absolute standards such as school means which are then compared with national norms, (2) analyzing trends in test scores for a given grade level over a period of years, (3) analyzing trends in test scores for a given cohort of students as they progress through a school, possibly comparing their performance to national normative data, and (4) applying various methods based on residuals from a regression analysis. The Dyer et al. (1969) method is seen to be a combination of the 3rd and 4th approaches.

As might be expected, a number of studies were conducted to determine the consistency of results under different methodologies and, not surprisingly, the evidence indicated that different

approaches tend to lead to the identification of different sets of schools as effective (see, e.g., Frechtling, 1982; Frederick and Clauset, 1985).

Some researchers have been critical of the methods used to identify effective schools (e.g., Rowan et al., 1983; Ralph & Fennessey, 1983; and Purkey & Smith, 1983). A common area of criticism has been the sole dependence on achievement data and the fact that sometimes only one subject area (e.g., reading) was considered. It was also noted by some critics that it may be inappropriate to label a school as effective if test scores for students at only one or two grades are examined.

Monitoring and Awarding School Performance

These criticisms concerning the rather limited nature of some of the early methods used to identify effective schools have had an effect on researchers charged with monitoring school performance and/or identifying the schools to receive some form of recognition. Thus we see the inclusion of a whole host of indicators such as student and teacher attendance, dropout rates, the percentage of students enrolled in "academic" courses, averages SAT scores, etc. Furthermore, SEIs based on student achievement typically include at least the two basic skills areas of reading and math and generally are available for all or at least most of the grades represented in a school. In a number of situations these grade-level and subject area SEIs are aggregated in some way to the school level and the resulting index is one of the main factors -- or the only factor -- in the determination of the schools to receive recognition or a cash award.

Stability of SEIs

The research on the stability of SEIs has focussed on a variety of issues. Dyer et al. (1969) assessed the stability of SEIs in five subject areas for 8th graders in a sample of 64 school systems (rather than schools) in New York state. In this investigation the authors demonstrated that SEIs based on random subsamples of children were quite stable, the stepped-up correlations between the pairs of SEIs ranged from .77 to .91 depending on the subject area of the output variable. The results of a similar study by Marco (1974) for a sample of third grade Title I students enrolled in 70 elementary schools in the Midwest were quite similar to the findings of Dyer et al. (1969).

Although these reliability coefficients are quite high, it is important to remember that they reflect the stability of residual achievement in one subject area for two random samples of children who were in the same grade. They provide no evidence, however, concerning the consistency of SEIs when factors such as grade level and subject matter are allowed to vary.

Mandeville and Anderson (1986) considerd the stability issue as it relates to cross-subject area and cross-grade SEIs. Cross-subject area SEIs (based on the same students) were reasonably stable (rs in the .60s to .70s) but cross-grade SEIs (based on different students) were very unstable (median r of .06 in reading and .13 in mathematics). Marco (1974) also considered the stability of SEIs for different groups of students based on ability (high and low) and found a correlation of .32.

Matthews, Soder, Ramey and Sanders (1981) studied the

stability of SEIs using the public schools in Seattle as their sample. They reported that the SEIs were quite unstable across grades (2-8), subject areas (reading and mathematics), and years (1978-79 and 1979-80). Year-to-year correlations of SEIs computed at the same grade level ranged from -.24 to .44, none of which were statistically significant because of the small number of schools involved.

Another of the relatively few studies to address the stability issue across years was conducted by Forsyth (1973) who used data for two successive graduating classes in 50 high schools. Outputs were twelfth grade test scores; inputs were the test scores for these students when they were ninth graders. For the nine subtests of the Iowa Tests of Educational Development (and the Composite) the correlations among the residuals for the two years ranged from .11 (Quantitative Thinking and Vocabulary) to .50 (Social Studies) with a median of .28.

Instruments

For a number of years, South Carolina has tested all students in the majority of the grades in the K-12 span. Criterion referenced tests (CRT) used as a part of the Basic Skills Assessment Program (BSAP) are administered each spring to all students in grades 1, 2, 3, 6, 8, and 11. Students in grades 4, 7, and 10 are tested, also in the spring, with the Comprehensive Tests of Basic Skills (CTBS). In addition, the Cognitive Skills Assessment Battery (CSAB) is administered at the beginning of the first grade as a school readiness test. The BSAP reading and mathematics tests are relatively short (36 and 30 multiple choice items respectively)

and provide, among other metrics, scale scores.

Methods

This study was limited to the elementary schools in South Carolina which contained grades one through four (and possibly additional grade levels) during both the 1984-85 and 1985-86 school years. During various of the data processing steps, however, students from other schools were included. For purposes of simplicity, preliminary processing will be described for the data from the 1984-85 school year only since those applied to the 1985-86 data were identical.

Student records for the Spring 1985 testing were matched with the corresponding test records for the most recent prior testing. That is, Spring 1985 CTBS test records for 4th graders were matched with their Spring 1984 BSAP records when they were in the 3rd grade; Spring 1985 BSAP records of 2nd and 3rd graders were matched with their corresponding Spring 1984 BSAP records when they were 1st and 2nd graders, respectively; and Spring 1985 BSAP records of 1st graders were matched with their CSAB records from Fall 1984. At this stage, the data files contained <u>all students</u> in South Carolina who had been tested and were successfully matched. Reading and math scores based on the Spring 1985 testing were to serve as the output variables for each of the four grade cohorts; prior year BSAP reading and math scores or fall CSAB raw score (depending on grade level) were used as student input variables; and the indication of eligiblity for free or reduced price lunches was to serve as a hard-to-change variable.

Once the matched data files were prepared, school-grade level

means were obtained for the two output variables and the input variable(s). During this aggregation step, the percentage of children eligible for free or reduced price lunches (LUNCH) was also obtained. (It is important to note that lunch data were aggregated from student records to the "grade within school" level to avoid aggregation error.) Finally, school-grade level aggregates based on fewer than 20 complete records were eliminated. (Generally speaking, 20 - 30 school-grade level records were eliminated for each of the data sets, but in each case more than 500 were retained.)

Next, regression analyses were conducted for each grade with each of the two output variables regressed onto all input variables and LUNCH. The residuals were divided by their estimated standard errors to produce what have been called "studentized residuals" and these were used as the SEIs.

When the above processing was completed for the 1984-85 data, SEIs were produced for 542 schools which contained the 1st grade; 535 with 2nd graders; 532 with 3rd graders; and 519 with 4th graders. For the 1985-86 data, the number of SEIs produced per grade were roughly the same. When these eight data sets were merged, only those schools with SEIs at all four grades for both years were retained. The resulting number of schools was 431. This reduction in n was due to the various grade organizational patterns of schools in South Carolina, the closing and opening of schools, etc.

Once a compatable data set was created, various aggregate indices such as those often used to make awards to schools were

created. Included were 1) a reading and math composite (RMCOMP) which was the average of the SEIs for the two subject areas and was computed for each grade, and 2) weighted (by sample size) averages (WAVEs) across grades for the reading, math and RMCOMP SEIs. These were computed for both years under study. The final step was to compute the correlations among the various SEIs.

Results

A summary of the results of the regression analyses for the two years is presented in Table 1. In terms of R^2 they are seen

---

Insert Table 1 about here

---

to be quite similar for the two years with the possible exception that the 1986 predictions appear to be more precise than those for 1985 at grade 4. Except for the one case noted (BSAPM was a significant predictor of CTBSR at the 4th grade in 1986 but not 1985) the significant ($p < .05$) predictor sets were consistent across years and the coefficients were quite similar in magnitude. The squared multiple correlations for predicting reading ranged from the .40s at grade 1 to the .70s at grade 4.

Predictions were less accurate in math than in reading, the R^2s ranging from about .3 (grade 1) to the .40s (grade 4). Rowan and Denk (1983) also reported better predictions in reading than math for the California Assessment Program data. This result led Mandeville and Anderson (1986) to conjecture that schools may potentially have more impact in mathematics than in reading. We also note that reading is a significant predictor of all output

variables whereas math is usually only relevant to math outputs. The finding that predictions are more accurate at the higher grades is well-known and caused Webster and Olson (1984) to conjecture that "at higher grade levels schools may have a lesser opportunity to show a differential effect on student achievement." (Webster & Olson, 1984, p. 7).

Although year-to-year stability is the focus of this study, the results on cross-subject area and cross-grade stability will be presented first as background information. Table 2 contains the correlations measuring the degree of relationship between the SEIs

---

Insert Table 2 about here

---

for reading and math. The correlations are seen to be very consistent across the two years and are quite large, especially in the 1st grade. Apparently adjusted reading and math performance for a given group of students in a school is reasonably similar. Cross-grade weighted averages in reading and math -- WAVEs -- are seen to be no more stable than the grade specific SEIs.

The other background data are the cross-grade rs which are presented in Table 3. Here the correlations are seen to be much

---

Insert Table 3 about here

---

smaller than those in Table 2 but are reasonably consistent for the two years. The major differences in the correlations for the two years are in the reading area where the SEIs for grades 1 and 3 are

significantly correlated with those for grade 4 for the 1985 data but not for the 1986 data. In both subject areas the correlations between the grade 1 and 3 and between the grade 2 and 3 SEIs were significant and similar in size for the two years. Grade 1 and 2 SEIs were also related in math but not in reading.

Although there are a few exceptions, generally speaking grade 4 SEIs are not related to those for the earlier grades. This relative independence of grade 4 SEIs may be related to the different output measure -- the CTBS -- used at this grade level. Each of the BSAP tests assesses student achievement on the same set of six reading and five math objectives developed for South Carolina so that they may be more curriculum valid than the CTBS. The final observation from Table 3 is that the correlations in math are somewhat larger than those in reading, especially for 1986.

The difference in the magnitude of the rs in Tables 2 and 3 might cause one to infer that when the <u>same</u> children are involved (Table 2) SEIs will tend to be moderately related but that when SEIs are based on <u>different</u> children, they are likely to be unrelated.

The year-to-year correlations of the SEIs, presented in Table 4, provide further evidence on this issue. First, the eight

---

Insert Table 4 about here

---

grade-specific rs are seen to be somewhat larger than those found by Forsyth (1973), which ranged from .11 to .50 (median = .28). Two factors which might account for this are that Forsyth's

analysis involved older students -- 12th graders -- and that his input data were gathered three years earlier, when the students were in grade 9.

We note that the year-to-year stability of reading SEIs is approximately equal to the year-to-year stability of the math SEIs for grades 1 - 3 but there is a definite decrease in stability across these three grades. That is, in grade 1 both correlations are in the .60s; in grade 2 they are in the .40s; and in grade 3 the correlations are in the .30s. The similarity of year-to-year stability for reading and math SEIs does not occur at grade 4 where the math SEIs appear to be more stable than those for reading (r = .50 vs. r = .36).

The SEIs based on averaging reading and math (RMCOMP) are of roughly the same stability as the separate reading or math SEIs, indicating that this approach does not produce an increase in year-to-year stability. As might be expected from the earlier results, aggregating across grades does not aid year-to-year stability either, as evidenced by the bottom row of Table 4. The "double composites" (across both subject areas and the four grades) correlate .45 year-to-year.

Considering the results presented in Tables 2 and 3 above, the correlations in Table 4 are surprisingly large. In grade one, for example, the year-to-year stability of the reading and math SEIs -- based on different children -- are in the .60s. Therefore, they are almost as large as the within-year correlations between reading and math SEIs -- based on the same children -- which are in the low .70s. Although the differences at the other three grades are

somewhat larger, the rationale that the low grade-to-grade correlations are simply due to fact that different children are involved is not supported.

In addition to addressing the issue of year-to-year stability at the same grade level (different students), we might also consider the year-to-year stability of SEIs based on the same students (at different grade levels). Although the data do not provide precise evidence here, most of the students on the 2nd grade 1986 data set were also on the 1st grade 1985 data set, and most on 3rd grade 1986 data set were on the 2nd grade 1985 data set, etc. Although it is true that the two data sets for the same grade-cohort do not contain exactly the same children, most of the transient students were eliminated from both data sets during the matching.

An important aspect of the analytic method used in this study is that the outputs for the first year, 1985, became inputs for 1986. Thus, if the 2nd graders in a school did particularly well in 1985 the input means to predict their 1986 performance as 3rd graders would be large and relatively high 1986 output means would be predicted. If their 1985 performance as 2nd graders was due to some school-wide factor, then the 3rd grade prediction is probably reasonable; to the extent that their exceptional 2nd grade performance reflected either random error variation or factors which, although consistently present at grade 2 of the school do not operate at grade 3, the prediction will most likely exceed their actual performance. By a parallel argument this form of the regression effect would produce 1986 predictions which are too low

for student groups who performed poorly the previous year for random or grade-specific reasons. These considerations lead to the prediction that if exceptional performance at a given grade level is due to random or grade-specific factors, the year-to-year correlations for grade-cohorts should be negative.

The year-to-year correlations for the two cohorts of students who were 2nd graders in 1986 (1st graders in 1985) and 3rd graders in 1986 (2nd graders in 1985) were small and nonsignificant, all being less than .10 in magnitude. However, for the cohort of 4th graders in 1986 (3rd graders in 1985), the correlations were negative and all were significant at the .01 level for one-tailed tests ($r = -.15$ for reading; $r = -.12$ for math; and $r = -.16$ for the RMCOMP).

Discussion

In this paper I have attempted to assess the stability of SEIs and extended the earlier work of Mandeville and Anderson (1986) to include year-to-year variation in addition to variation across subject areas and grades. There are a number of limitations with SEIs which are based on school-grade means on two achievement tests and the author readily acknowledges them. It would clearly be more appropriate to: 1) include additional achievement measures in other subject areas, 2) include measures outside the achievement domain, 3) consider more complex operational definitions than the school-grade level mean. This research, however, did not have the objective of identifying more appropriate procedures for use in school recognition and reward programs.

Before proceeding, it should be noted that the first two

points above may not be as relevant in this research as they would be for research conducted in other states. This is because of the statewide emphasis on the basic skills -- particularly reading and math -- and the fact that the BSAP tests were developed based on statewide objectives. The legislature has mandated that achievement in these two subjects as measured by the BSAP should be priority areas for all schools.

One aspect of the interpretation of these results is to consider the evidence provided regarding the consistency of SEIs as they operate throughout a school and across years. The results on this issue were:

1. Reading and math SEIs for the same students in the same grade correlate in the .60s to .70s.
2. Year-to-year SEIs for different student groups at the same grade level correlate in the .30 to .60 range.
3. Correlations of cross-grade SEIs never exceed .20 and are often nonsignificant.
4. The prediction that SEIs for the same grade cohorts would be negatively correlated was upheld for the 4th grade (1986) cohort only; in every case, the correlations were very small, however.

In finding #1 above, the students and teachers are "controlled" in the sense that the same students and teachers are involved; in finding #2 the students differ but, except for changes in teacher assignments and other "mortality" factors, it is likely that the same teachers are involved; finding #3 reflects different students and teachers; and, finally, in finding #4 the students were largely

the same, but the teachers differed. The constant factor associated with nontrivial correlations in findings #1 and #2 above is the grade level and, presumably, the group of teachers who were responsible for instruction at that grade level.

To put it another way, about 40% to 50% of the variation in math performance for a group of students in a given grade and school can be predicted based on their reading performance (and visa versa for predicting reading from math). If we use the performance of students who were enrolled in a given grade during the previous year to predict the performance of students at the same grade for the current year, between 12% and 42% of the variation will be accounted for depending on the grade and subject area. Finally, if we wish to predict from one grade to another, whether the same or different students are involved, the best we can hope to do is to predict about 3% of the variation. The evidence, although not overwhelming, strongly suggests that factors operating at the grade level dominate whatever overall school factors affect student achievement.

Other findings which were not specifically related to the above issue were:

1. Reading and math SEIs are of about the same year-to-year stability for grades 1 - 3 where the BSAP tests are the output.
2. Year-to-year stability decreases from the .60s in grade 1 through the .40s in grade 2 to the .30s in grade 3.
3. The results at grade 4, where the CTBS is the output, do not agree with those for grades 1 - 3; reading SEIs

continue to trend downward in stability but the math SEIs are somewhat more stable than those for grades 2 and 3. Although I do not feel at all confident about them, the following are offerred as post-hoc speculations about these results. Possibly, they will provide some direction for future research. Although the SEIs are based on aggregates, interpretive statements at the student level are not erroneous and will be made for simplicity.

One possible explanation of #1 above is that both the reading and math SEIs derived from the BSAP tests largely reflect the children's ability to read. The correlations between reading and math SEIs (see Table 2) and the fact that BSAPR was always the best single predictor of BSAPM (related to the results in Table 1) supports this view.

Regarding point #2, it is probably true that the primary objective in the 1st grade is to teach children to read. It may be that given this narrow focus, "teachers effects" are more consistent at grade 1 than at higher grades. If other subject areas receive progressively more treatment in grades 2 and 3, the consistency of teacher effects, and therefore school effects, may diminish. It is also possible that the decrease in stability is an artifact of the methodology.

An explanation for #3 may rest on the differences between the BSAP math test, developed specifically to assess student performance on the South Carolina math objectives, and the more broadly based CTBS math test. Whereas all 3rd grade students are probably instructed with the BSAPM objectives in mind (students can

actually fail this test), it is possible that considerable variation exists in the "fit" of 4th grade math textbooks to the objectives measures by CTBSM. A related explanation would be that some of the more advanced topics included in the Concepts and Applications Subtest of the CTBSM (e.g., measurement and geometry), although included in the 4th grade curricula for most or all schools, are routinely not taught in some schools for lack of time.

References

Bandy, I. (1985, February). State education performance indicators. Paper prepared for the Steering Committee of the Counsel of Chief State School Officers, Washington, DC.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., Mood, A. M., Winfield, F. D., & York, R. L. (1966). Equality of educational opportunity. Washington, D. C.: U. S. Department of Health, Education and Welfare.

Dyer, H. S. (1970). Toward objective criteria of professional accountability in the schools of New York City. Phi Delta Kappan, 52, 206-211.

Dyer, H. S., Linn, R. L., & Patton, M. J. (1969). A comparison of four methods of obtaining discrepancy scores based on observed and predicted school system means on achievement tests. American Educational Research Journal, 6, 591-605.

Forsyth, R. A. (1973). Some empirical results related to the stability of performance indicators in Dyer's student change model of an educational system. Journal of Educational Measurement, 10, 7-12.

Frechtling, J. A. (1982). Alternative methods for determining effectiveness: Convergence and divergence. Paper presented at the annual meeting of the American Educational Research Association, New York.

Frederick, J. M. & Clauset, K. H. (1985). A comparison of the major algorithms for measuring school effectiveness. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Good, T. L. & Brophy, J. E. (1986). School Effects. In M. C. Whittrock (Ed.), Handbook of Research on Teaching (3rd ed., pp. 570-604). New York: Macmillan.

Hilton, T. L., & Patrick, C. (1970). Cross-sectional versus longitudinal data: An empirical comparison of mean differences in academic growth. Journal of Educational Measurement, 7, 15-24.

Marco, G. (1974). A comparison of selected school effectiveness measures based on longitudinal data. Journal of Educational Measurement, 11, 225-234.

Mandeville, G. K. & Anderson, L. W. (1986). An study of the stability of school effective measures across grades and subject areas. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Matthews, T. A., Soder, J. B., Ramey, M. C. & Sanders, G. H. (1981). Use of district test scores to compare the academic effectiveness of schools. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

O'Connor, E. F. (1972). Extending classical test theory to the measurement of change. Review of Educational Research, 42, 73-97.

Purkey, S. C. & Smith, M. S. (1982). Too soon to cheer? Synthesis of research on effective schools. Educational Leadership, 40, 64-69.

Purkey, S. C. & Smith, M. S. (1983). Effective schools: A review. Elementary School Journal, 83, 427-452.

Ralph, J. H. & Fennessey, J. (1983). Science or reform: Some questions about the effective schools model. Phi Delta Kappan, 64, 689-702.

Rowan, B., Bossart, S. T., & Dwyer, D. C. (1983). Research on effective schools: A cautionary note. Educational Researcher, 12, 24-31.

Rowan, B. & Denk, C. E. (1983). Modelling the academic performance of schools using longitudinal data: An analysis of school effectiveness measures and school and principal effects on school-level achievement. San Francisco, CA: Far West Laboratory for Educational Research and Development.

Webster, W. J. & Olson, G. H. (1984). An empirical approach to identifying effective schools. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Wynne, E. A. (1984). School award programs: Evaluation as a component of incentive systems. Educational Evaluation and Policy Analysis, 6, 85-93.

Table 1
Significant Predictors and Squared Multiple Correlations
By Output Variable, Grade and Year

Grade	Output	Significant Predictors	1985 N	1985 R^2	1986 N	1986 R^2
1	BSAPR	CSAB LUNCH	542	.46	550	.41
	BSAPM	CSAB LUNCH		.31		.27
2	BSAPR	BSAPR LUNCH	535	.67	529	.72
	BSAPM	BSAPR BSAPM LUNCH		.45		.47
3	BSAPR	BSAPR LUNCH	532	.65	534	.64
	BSAPM	BSAPR BSAPM LUNCH		.34		.36
4	CTBSR	BSAPR BSAPMaLUNCH	519	.72	526	.78
	CTBSM	BSAPR BSAPM LUNCH		.40		.49

Note. To be included as a "significant predictor", A regression coefficient was significant ($p < .05$). BSAPR and BSAPM refer to the reading and mathematics subtests of the BSAP, respectively, and CTBSR and CTBSM have the same meaning for the CTBS.
a significant for 1986 only.

Table 2
Correlations Between Reading and Mathematics SEIs

Grade	1985	1986
1	.73	.74
2	.61	.62
3	.61	.62
4	.59	.63
WAVE	.68	.72

Table 3
Pearson Correlations Among Grade-Specific SEIs
By Subject Area and Year

Grades	Reading 1985	Reading 1986	Math 1985	Math 1986
1 & 2	.02	.00	.19**	.14**
1 & 3	.10*	.13**	.16**	.16**
1 & 4	.12*	.04	.06	.09
2 & 3	.11*	.18**	.18**	.19**
2 & 4	.04	.00	.13**	.11*
3 & 4	.12*	-.02	.09	.00
Median	.105	.02	.09	.125

Note: *$p < .05$. **$p < .01$.

Table 4
Cross-year Correlations of Various SEIs
(1985 vs. 1986)

Grade	Reading	Math	RMCOMP
1	.65	.60	.66
2	.42	.46	.47
3	.39	.34	.39
4	.36	.50	.45
Wt. Ave.	.41	.46	.45

Note: $p < .001$ for all correlations.